DOCUMENT RESUME

ED 220 484 TM 820 497

AUTHOR	Nyberg, Verner R.; Nyberg, Adell M.
TITLE	Technical Report on Alberta Essay Scales: Models.
INSTITUTION	Alberta Univ., Edmonton. Faculty of Education.
REPORT NO	ISBN-0-88864-948-7
PUB DATE	82
NOTE	27p.; For a related document, see TM 820 496.
AVAILABLE FROM	Faculty of Education, University of Alberta, 4-116 Education North, Edmonton, Alberta, Canada T6G 2G5 ($8.00).
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Academic Standards; *Expository Writing; Foreign Countries; Grade 12; *Grading; High Schools; Holistic Evaluation; Rating Scales; Scoring Formulas; *Student Writing Models; *Test Reliability; *Test Validity; *Writing Evaluation; Writing Skills
IDENTIFIERS	Alberta; *Alberta Essay Scales

ABSTRACT

The supplementary information on "Alberta Essay Scales: Models" presented here includes similar models to employ in grading essays, the background and development of the scales, and the rationale for developing two scales of English mechanics and style/content. A standard is presented for evaluating current writing achievement by comparison with compositions at the 12th grade level written in 1964. Two scales are found necessary because a low correlation between mechanics and style/content is evident. The grading procedures employed in 1964 and present changes are discussed. A holistic approach to grading by total impression is presented as more efficient than atomistic grading by summing scores on several variables. Technical considerations include the methods for determining reliability and content validity, the scoring variables, and means as true scores for each essay. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

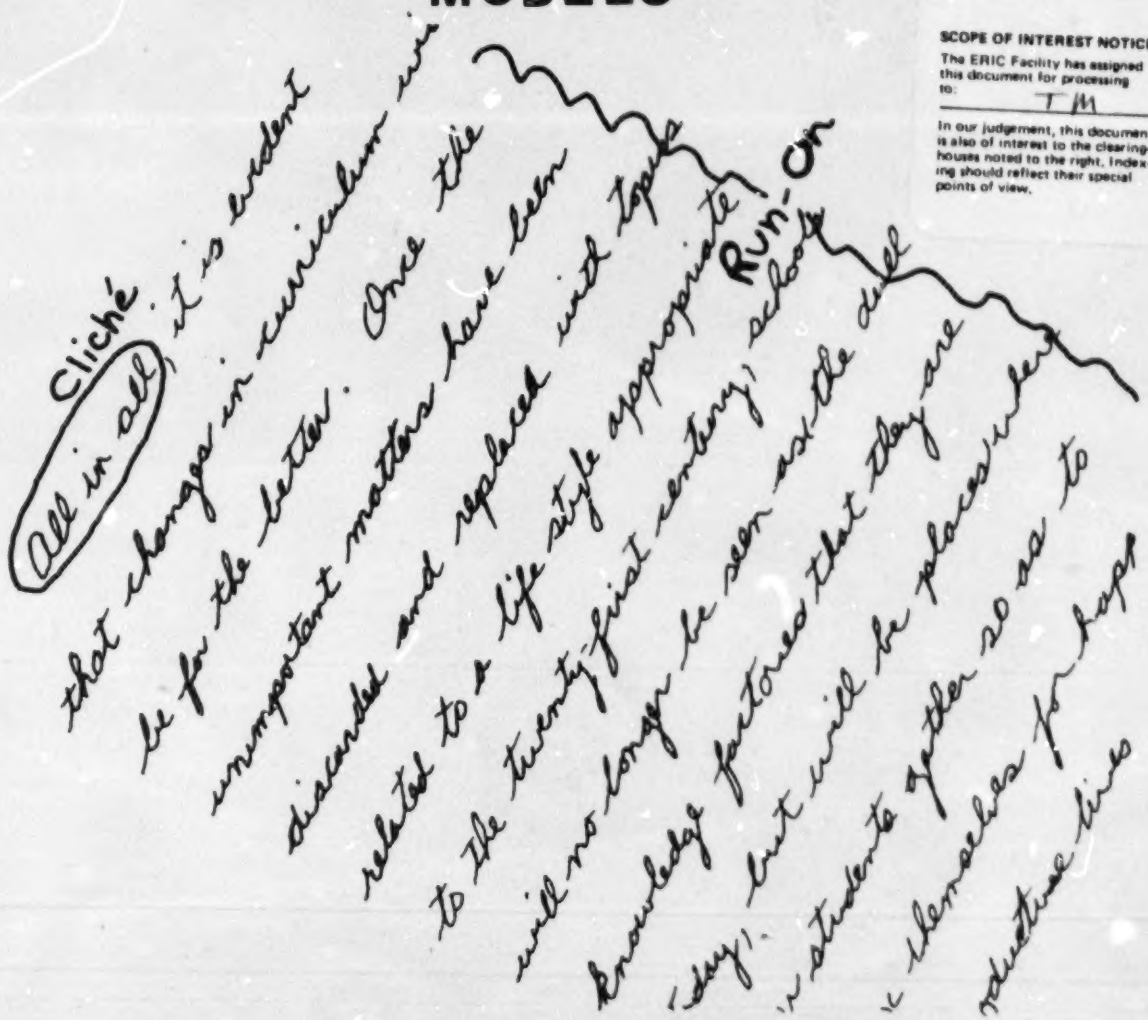
ED220484

TECHNICAL REPORT ON ALBERTA ESSAY SCALES: MODELS

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: TM CS

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.


Cliché
All in all, it is evident
that changes in curriculum are
be for the better. Once the
unimportant matters have been
discarded and replaced with topics
related to a life style appropriate
to the twenty-first century, schools
will no longer be seen as the dull
knowledge factories that they are
today. But will be places where
students gather so as to
themselves for happy
productive lives
Run-on

TM 820 497

VERNER R. NYBERG AND ADELL M. NYBERG

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

S. Ford

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TECHNICAL REPORT ON ALBERTA ESSAY SCALES: MODELS

VERNER R. NYBERG
THE UNIVERSITY OF ALBERTA

AND

ADELL M. NYBERG
SPRUCE GROVE COMPOSITE HIGH SCHOOL

1982

© FACULTY OF EDUCATION
UNIVERSITY OF ALBERTA
ISBN 0-88864-948-7

FOREWORD AND ACKNOWLEDGMENT

This technical report is intended to accompany *Alberta Essay Scales: Models* (Nyberg, V. and Nyberg A., 1982), and provides supplementary information. The booklet containing the models is sufficient for teachers who wish to employ the scales for grading essays. The technical report provides information on similar models, how the scales were developed, reliability and validity of the scales, and the rationale for developing two scales instead of one.

The authors gratefully acknowledge the assistance provided by Alberta Education in duplicating and distributing the original (197/) version of the scales.

TABLE OF CONTENTS

FOREWORD AND ACKNOWLEDGMENT	iii
TECHNICAL REPORT ON ALBERTA ESSAY SCALES: MODELS	1
Background	1
Development of the Scales	2
Grading Procedures Employed in 1964	4
GENERAL DIRECTIONS FOR USE OF THE SCALES	5
TECHNICAL CONSIDERATIONS	6
Reliability of Scoring Through Use of the Scales	6
Validity	9
Need for Two Scales	9
Means as "True" Scores	11
Deviations From 1964 Procedure	11
CONCLUDING STATEMENT	14
REFERENCES	15
APPENDIX A	17
Form I	18
Form II	19
Form III	20
Form IV	21

TECHNICAL REPORT ON ALBERTA ESSAY SCALES: MODELS

BACKGROUND

Evaluation of written composition through the use of models is not a new technique. The Hillegas Composition Scale (Hillegas, 1912) was one of the first and perhaps the best known scale to be used in North America. In the early 1920's other scales, similar to the one devised by Hillegas, appeared on the scene. Some of the more familiar ones included the Willing Scale (Willing, 1922), the Van Wagenen English Composition Scale (Van Wagenen, 1923), and Hudelson's Typical Composition Ability Scale (Hudelson, 1923). At least some of these scales seem to have been employed until the 1940's.

In recent years other models have been developed and some have become well known. Perhaps the best example is the set of models published by Educational Testing Service (1957). A scale by Bader and Steinhoff (1958) and one entitled *A scale for Evaluation of High School Student Essays* (California Association of Teachers of English, 1960) are less well known, as are those described in the *Illinois English Bulletin* (1965), and in the *Kentucky English Bulletin* (1965).

The various essay scales all employ essentially the same technique. A set of typical essays, or models, usually consisting of from five to ten compositions, provides examples varying in quality from poor to excellent. Each model essay is given a number representing its achievement level. The numbers may indicate grade levels, percentiles, or other scale values. If a teacher wishes to evaluate a group of essays, each one is, by inspection, matched with a model essay of similar quality, and the value corresponding to the model is then placed on the essay being graded.

As might be expected, very little statistical analysis was done on the early scales, and such research that was performed was often of questionable

quality. Some of the later scales, however, have been subjected to rather rigorous analyses.

The purpose in making another set of models available for use is not to supplant other models but to meet needs specific to teachers of twelfth grade academic English. Some of these specific needs are:

a) for teachers, a scale based on compositions written by grade twelve students some years ago; that is, spring, 1964.

b) for researchers in various school systems, a scale by means of which longitudinal studies of achievement in composition can be made.

The second purpose was considered to be especially important in that there was a need for a procedure for documenting (or refuting) claims that writing skills have declined. The writing samples shown in *Alberta Essay Scales: Models* provide a means for checking writing skills against standards prevailing in 1964.

Development of the Scales

Two composition scales were developed in this study. One deals with English mechanics such as spelling and punctuation, and the other deals with style and content which includes factors such as richness of vocabulary and literary devices, as well as actual substance. Two scales instead of one are considered defensible for two reasons. From a logical point of view there seems little similarity between accuracy in spelling, punctuation, and grammar, as opposed to how a composition is organized, what level of diction is employed, and what material is actually prescnted. Usually both of these broad aspects are considered when compositions are scored, however, it appears that schievement scores in these two fields are almost always combined into a single grade. From a statistical viewpoint the use of two grades instead of one seems necessary. The data from which the scales were produced (these data are described later)

indicate a low correlation between mechanics and style-content, hence it is impossible to interpret a single score derived from the two fields. It is obvious, however, that teachers who do not agree with this argument cannot be prevented from combining the two scores in any manner they wish.

The two scales are derived from data based on English 30 (grade 12 academic English) Department of Education examinations administered in Alberta in June, 1964. In July the compositions submitted by 8,020 candidates (Nichols, 1966) were graded by two committees of English teachers under the direction of the Examinations Branch of the Alberta Department of Education. The practice was to appoint one committee to grade for proficiency in the mechanics of English, and another committee to grade for style and content. According to Nyberg (1966), the committees consisted of twenty-seven and twenty-one raters respectively, and in charge of each was a chairman whose responsibilities included spot checking each rater's work to see that the instructions were being followed.

Great care was taken to make the ratings as uniform and as reliable as possible. Each essay was rated on specified variables, six for mechanics and sixteen for style and content. Each variable was carefully defined before the grading commenced. The rating forms used and the written descriptions of the variables are shown in Appendix A.

A random sample of one hundred and three essays was selected from the population of essays written on the topic "Advertising," the most popular of the two topics specified in the test booklet employed in 1964. Any essay drawn by way of the sampling procedure that was well below the specified length (400-500 words) was not retained for further study.

Each essay in the sample was scored many times, but all grades were entered on separate scoring forms. No marks of any kind were entered on the students'

essay booklets, therefore, no teacher knew how any essay had fared at the hands of the others.

From the 1964 data, two essay scales were constructed in 1977. This was done by computing stanines for the mechanics and style-content groups and then selecting representative essays from stanines two, four, six, and eight. Each selection was made on the basis of two criteria; first, the standard deviation of the scores of the essay selected was relatively low, and second, the mean score assigned to the essay was near the middle of the stanine interval.

Grading Procedures Employed in 1964

The English essays connected with the grade twelve examinations were scored by reference to two forms that specified the variables to be considered and the weighting to be attached to each. One form was for grading mechanics and the other was for style and content. These forms are shown in Appendix A. In order to avoid differences in interpretation of the variables two other forms were provided that described each variable in some detail. Form A, for mechanics, used a deduction system. Each essay was first awarded 125 marks, then marks were taken away for each error encountered. If there were many errors the full 125 marks might be lost. Form B, the style-content form, used a bonus system. Marks were awarded on each of sixteen variables signifying the degree of success achieved. The raw score awarded to an essay was determined by adding 125 marks to the total for Form B, then subtracting the deductions accumulated on Form A. The raw scores were then transformed into scores on a scale ranging from zero to one hundred. These transformed scores are of no particular interest for this study; instead, stanines are deemed significant.

GENERAL DIRECTIONS FOR USE OF THE SCALES

The essential task, when employing the models for grading, is to match each essay with one of the model essays. In order to achieve as much reliability and validity as possible, it is necessary to have a clear idea of the variables that were taken into account when the models were graded originally.

The variables taken into consideration in 1964, as well as the weighting assigned to each, are shown in Appendix A, along with forms that explained the meaning of each variable in some detail. It should be noted that the two variables Plan and Relationship of Plan to Essay are not to be considered when essays are compared to models. The reasons for this are given on page 13 of this report. The Alberta Essay Scales should not be used unless the rater is thoroughly familiar with the variables. Familiarity is best gained through practice, and special provision has been made for this. It is suggested that prospective raters practice their skills on both sets of models. For instance, the raters should first rate the mechanics models to see if they agree with the stanine levels given. Then they should rate the style-content models for mechanics and compare their results with the ratings given on page 34 of the Alberta Essay Scales: Models booklet. Next the raters should score both sets of models for style content. Page 35 of the essay scales booklet shows the ratings that were given to the mechanics models when rated for style-content.

The model essays represent stanines 2, 4, 6, and 8. If an essay to be graded is judged poorer than stanine 2 it would be given a stanine of 1. An essay might be considered better than a 2 but inferior to a 4, therefore it would be awarded a 3. Similarly, stanines 5 and 7 would be given when the essays did not fit the models corresponding to 4, 6, or 8. A value of 9 would be given an essay judged superior to the model labelled 8. It is possible, therefore, through the use of only 4 model essays to identify nine levels of achievement.

In order to assist teachers in using the models, a table providing the mean score for each of the variables is shown below each sample essay. It should be noted that the stanine values of the variables sometimes differ from the stanine value of the composition as a whole. This, of course, is to be expected, and raters should take these differences into account.

TECHNICAL CONSIDERATIONS

Reliability of Scoring Through Use of the Scales

Information on reliability of ratings based on composition models is scarce. Darsie (1912) claimed a reliability of .88 for the Willing Scale on the basis of ratings by two readers. This value was probably inflated because of the fact that the grades awarded by the first rater were entered on the papers where they could be seen by the second rater. The scales published by Educational Testing Services (1957) were checked for reliability, and values from .62 to .70 were reported. These reliability estimates were based on two sets of readings by the same reader. Reliabilities of .77 for mechanics and .60 for style-content were reported by Nyberg (1966) in the study from which the models for the Alberta Essay Scales originated.

The reliability of the scoring when the *Alberta Essay Scales: Models* was used was investigated in a study by Nyberg (1977). Each of four teachers was asked to select a class of grade twelve academic English students and to give them a composition test. The test consisted of writing an essay of approximately 200 words on the topic "The Ideal School." Eighty-three students in all took the test. Six experienced teachers of English then rated all the essays independently using *Alberta Essay Scales: Models*. No special instructions beyond those given in the scale booklets were provided. Each of the essays was graded for both mechanics and style-content.

The next step was to determine the correlation between each pair of

markers. The reliability of the scoring procedures was then estimated by determining the mean of the correlations between pairs of markers. In addition to this, means and standard deviations were computed for the scores awarded by each of the readers. These values were used to check whether some of the markers were unduly easy or difficult or if some tended to award more extreme scores than others. Table 1 shows the correlations between mechanics scores awarded by the six readers, and Table 2 shows the means and standard deviations. Table 3 and Table 4 show corresponding values for the style-content scores.

Table 1

Correlations for Mechanics Scores Awarded by Six Raters

Rater Number	Rater Number 2	3	4	5	6
1	.949	.836	.743	.880	.830
2		.823	.672	.821	.809
3			.592	.738	.689
4				.837	.698
5					.820

Table 2

Means and Standard Deviations of Mechanics Scores

Rater Number	Mean	Standard Deviation
1	5.096	1.588
2	4.964	1.594
3	5.193	1.548
4	5.012	1.654
5	5.048	1.575
6	5.277	1.688

Table 3

Correlations Between Style-Content Scores Awarded by Six Raters

	Rater Number				
Rater Number	2	3	4	5	6
1	.952	.735	.768	.903	.844
2		.704	.757	.866	.815
3			.624	.671	.609
4				.822	.687
5					.820

Table 4

Means and Standard Deviations of Style-Content Scores

Rater Number	Mean	Standard Deviation
1	5.759	1.359
2	5.699	1.462
3	6.470	1.320
4	5.181	1.300
5	5.470	1.329
6	5.940	1.563

The reliability of the scoring for mechanics was calculated to be 0.782. The means appeared to be quite close, but t-tests indicated that differences were significant (p=.05) between raters 1 and 2, 2 and 3, 2 and 6, and 5 and 6. Differences between pairs of standard deviations were not significant (p=.05) in all cases.

For the style-content aspect of the essays the reliability of scoring was found to be 0.772. Differences between pairs of means were significant (p=.05) in every case except for readers 1 and 2, and readers 1 and 6. None of the differences between standard deviations were significant at the five percent level.

The reliability estimates obtained for the two scales compared favorably with the values reported by Educational Testing Services (1957) and by Nyberg (1966). Regarding the Alberta Essay Scales it was concluded that "the two sets of models can be used by teachers for grading twelfth grade English compositions." It would appear, however, that mathematical corrections might have to be applied to style-content scores awarded by different markers in order to reduce differences in standards.

Validity

No evidence of empirical validity has been gathered with respect to the Alberta Essay Scales, however, a degree of content validity can be claimed. The variables listed on the 1964 scoring forms are essentially those reported by Buxton (Braddock, Lloyd-Jones and Schoer, 1963) as being developed by two English specialists in Alberta.

Need For Two Scales

The decision to develop two scales, one for mechanics of English and one for style and content, was based on some rather convincing empirical

evidence. A study by Nyberg (1968) revealed that combining the two sets of variables caused an undue weighting to be given to mechanics because of the relatively low variances of the style-content variables. Table 5

Table 5

Statistics Related to Essay Scoring Variables

Variables	Max. Score Possible	Mean Score Awarded	S.D. of Scores	Contribution to Total Score (%) Expected	Actual
Mechanics					
Spelling	39	20.9	10.2	11.1	27.6
Punctuation	26	15.7	3.3	7.4	8.5
Word Usage	24	16.3	2.9	6.9	7.7
Grammar	33	24.2	3.9	9.5	12.2
Sentence Errors	39	29.4	4.5	11.1	14.5
Form	14	9.7	2.2	4.0	3.3
Total Mechanics	175*	116.2		50.0	73.8
Style-Content					
Significance	10	6.0	.6	4.0	1.6
Relevance	10	6.3	.6	4.0	1.6
Originality	5	1.8	.7	2.0	1.4
Plan	5	3.0	.6	2.0	1.0
Relation of plan to essay	5	3.3	.6	2.0	1.0
Introduction	5	2.6	.6	2.0	1.2
Order	5	2.7	.4	2.0	1.2
Emphasis	5	2.2	.5	2.0	1.3
Conclusion	5	2.3	.6	2.0	1.3
Vividness of words	10	5.8	.8	4.0	2.3
Figures of speech	5	2.4	.4	2.0	1.0
Vocabulary	5	2.5	.4	2.0	1.1
Sentence Structure	15	8.1	1.0	6.0	3.0
Sentence Beginnings	5	2.8	.3	2.0	.8
Economy	15	8.0	1.0	6.0	3.0
Total impression	15	8.0	1.2	6.0	3.4
Total Style-Content	125	67.8		50.0	26.2

*Total scores for mechanics were not permitted to exceed 125.

indicates that the mechanics variables' contribution to the total score was 73.8% while the style-content variables contributed only 26.2%. The intention of the Department of Education was to have each set of variables contribute 50%. It would be possible to develop a mathematical procedure to weight the two sets equally, but this would make the use of the scales

unduly complex. Also, further evidence strengthened the argument for two scales. Table 6 shows the correlations between all pairs of variables.

Before one attempts to interpret the table it is necessary to bear in mind that mechanics scores are deductions for errors, therefore negative correlations are expected between a mechanics variable and a style-content variable. It is immediately evident that the mechanics variables are moderately intercorrelated, the style-content variables very highly intercorrelated, but that correlations between the two sets is rather low. It was found, also, that the mean scores awarded by the 27 mechanics raters, and the mean scores awarded by the 21 style-content raters correlated only .47 over the sample of 103 essays. This relationship is perhaps no stronger than what would be found by intercorrelating such divergent measures as high school physics scores and foreign language scores.

Means as "True" Scores

The mean scores computed for each essay for mechanics and for style-content are considered to be "true" score estimates as defined in classical test theory. The correlation of .47 between the two sets of means, referred to in the previous section, is related to the degree of overlap between the constructs called "mechanics of English"and "style-content." In other words, the correlation of .47 is low not because of the unreliability of scoring, but because the two constructs are quite different.

Deviations From 1964 Procedure

Deviations from the 1964 procedure with respect to mechanics are minor. All of the variables are retained but the differential weighting (three marks deducted for some errors, and two for others) is abandoned. Instead, the suggested procedure is that the frequency of mechanical errors should be noted

TABLE 6

Matrix of Correlations Between All Pairs of Variables

Variable Number																						Variable Number
2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
.556	.523	.539	.537	.426	.662	-.277	-.313	-.154	-.277	-.320	-.241	-.360	-.262	-.269	-.367	-.238	-.322	-.405	-.381	-.425	-.340	-.366	1
	.505	.484	.589	.345	.647	-.164	-.233	-.019	-.244	-.275	-.199	-.279	-.104	-.134	-.241	-.127	-.187	-.267	-.206	-.322	-.227	-.242	2
		.605	.669	.323	.674	-.310	-.306	-.150	-.144	-.150	-.205	-.352	-.221	-.232	-.303	-.188	-.267	-.319	-.297	-.382	-.305	-.308	3
			.783	.195	.691	-.506	-.508	-.422	-.204	-.274	-.404	-.598	-.494	-.508	-.541	-.469	-.557	-.596	-.598	-.623	-.575	-.579	4
				.327	.734	-.504	-.531	-.286	-.353	-.390	-.424	-.594	-.427	-.396	-.515	-.402	-.484	-.550	-.496	-.596	-.517	-.548	5
					.420	.023	-.019	-.255	-.326	-.227	-.006	.002	.139	.056	.051	.160	.035	.008	.018	-.048	.078	.016	6
						-.396	-.434	-.208	-.333	-.370	-.336	-.497	-.337	-.351	-.456	-.313	-.419	-.502	-.465	-.548	-.445	-.473	7
							.927	.800	.206	.343	.797	.914	.902	.809	.913	.841	.904	.912	.847	.908	.940	.943	8
								.724	.254	.378	.768	.891	.849	.778	.904	.779	.863	.899	.808	.907	.910	.919	9
									.051	.200	.690	.794	.927	.785	.806	.861	.814	.813	.796	.791	.880	.837	10
										.883	.148	.201	.095	.062	.231	.240	.188	.198	.206	.245	.182	.239	11
											.241	.331	.251	.188	.342	.297	.336	.333	.342	.371	.325	.389	12
												.817	.772	.692	.830	.698	.755	.772	.734	.756	.788	.799	13
													.901	.818	.904	.823	.890	.935	.855	.929	.935	.943	14
														.860	.898	.878	.870	.906	.854	.882	.941	.930	15
															.808	.715	.798	.805	.769	.797	.839	.824	16
																.837	.905	.933	.884	.923	.945	.952	17
																	.881	.865	.855	.841	.891	.880	18
																		.921	.913	.901	.937	.935	19
																			.895	.943	.954	.953	20
																				.887	.912	.909	21
																					.960	.951	22
																						.974	23

List of Variables for Table 2

1. Spelling
2. Punctuation
3. Word Usage
4. Grammar
5. Sentence Errors
6. Form
7. Total Mechanics
8. Significance
9. Relevance
10. Originality
11. Plan
12. Relationship of Plan to Essay
13. Introduction
14. Order
15. Emphasis
16. Conclusion
17. Exactness
18. Figures of Speech
19. Vocabulary
20. Sentence Structure
21. Sentence Beginnings
22. Economy
23. Total Impression
24. Total Style-Content

in essays to be scored and matched with a model having a similar error frequency. This procedure is much simpler to use in that it is not necessary, for scoring purposes, to decide exactly what kind of error is present. For example, it is often difficult to decide whether an error involves grammar or word usage or perhaps punctuation, but there is no doubt that an error is present. It should be noted, also, that the procedure advocated by Buxton (Braddock, Lloyd-Jones and Schoer, 1963) did not suggest different penalties for different types of errors. Weighting the errors equally causes only minor differences in ranking of the essays. This is because the types of mechanical errors are correlated.

Regarding style-content, the variables entitled Plan, and Relationship of Plan to Essay, are no longer included. Table 1 shows that the contribution of these variables was only one percent each, and Table 2 shows that these two variables (numbers 11 and 12) have low correlations with all of the other style-content variables. The most compelling reason, however, is than an astute student often writes the essay first and constructs the plan afterwards. He or she thus expects to garner the five marks allotted to Relationship of Plan to Essay.

A further deviation is with respect to the method of scoring. It is suggested that the variable named Total Impression be used for grading. This is advocated because of the high correlation (.97) between the Total Impression score and the Total Score (see Table 2). This correlation indicates that the total impression of an essay is very nearly the sole determiner of the score that is awarded for style-content. Grading by total impression is faster than grading by summing subscores on several variables, therefore the total impression method is more efficient. This procedure assumes, however, that raters are aware of the variables to be taken into account.

One might argue that there are two ways of explaining the high correlation. First, it is possible that a reader forms an impression of an essay as it is read and then awards subscores based on that total impression. The other possibility is that a rater subconsciously totals the subscores and then awards a total impression score that agrees with the subconscious total. The second view seems untenable, as it does not account for the high correlations between all pairs of style-content variables. It is therefore assumed that the first possibility best represents the situation.

CONCLUDING STATEMENT

As stated earlier (page 2), the purpose in constructing the Alberta Essay Scales was to meet needs specific to teachers of Grade 12 academic English. This purpose has been achieved. It has been demonstrated that models based on compositions written by Alberta students in English in 1964 can be used to grade writing done in 1977 with a reasonable degree of reliability. In addition, researchers who wish to make longitudinal comparisons of achievement in composition writing can accomplish this through use of the models.

As a result of the development of the Alberta Essay Scales it became evident that written composition should be graded so as to produce separate scores for mechanics and style-content. It was demonstrated also that the "holistic" approach (grading by total impression") was more efficient than the "atomistic" approach (grading by summing scores on several variables), at least for grading essays written by twelfth grade academic English students.

REFERENCES

Bader, Arno L. & Steinhoff, William R. "Rating and Analysis of Student Themes." A Guide for Evaluating Student Composition, ed. Sister M. Judine. Champaign, Ill.: National Council of Teachers of English, 1963, pp. 113-119.

Braddock, Richard, Lloyd-Jones, Richard, and Schoer, Lowell. Research in Written Composition. Champaign, Ill.: National Council of Teachers of English, 1963, pp. 58-70.

California Association of Teachers of English. A Scale for Evaluation of High School Student Essays. Champaign, Ill.: National Council of Teachers of English, 1960.

Darsie, Marvin L. "The Reliability of Judgments Based on the Willing Composition Scale." Journal of Educational Research, 5, 1922, pp. 89-90.

Educational Testing Service. Cooperative Sequential Tests of Educational Progress Handbook for Essay Tests Level 2. Princeton, N.J.: Educational Testing Service, 1957.

Hillegas, M. B. A Scale for the Measurement of Quality in English for Young People. New York: Bureau of Publications, Teachers' College, Columbia University, 1912.

Hudelson, Earl. Hudelson English Composition Scale. New York: World Book Company, 1923.

Illinois English Bulletin. "Evaluating Twelfth Grade Themes." A Guide for Evaluating Student Composition, ed. Sister M. Judine. Champaign, Ill.: National Council of Teachers of English, 1965, pp. 130-137.

Kentucky English Bulletin. "Two Types of Grading." A Guide for Evaluating Student Composition, ed. Sister M. Judine. Champaign, Ill.: National Council of Teachers of English, 1965, pp. 138-146.

Nichols, S. T. "Examinations Branch." Sixtieth Annual Report of the Department of Education of the Province of Alberta 1965. Edmonton, Alta.: Queen's Printer, 1966, pp. 77-79.

Nyberg, Adell M. "A Study of the Reliability of an Alberta Composition Scale." Unpublished paper presented at a graduate seminar, University of Alberta, 1977.

Nyberg, Verner R. "A Factor Analytic Study of Essay Gradings." Unpublished Ed.D. dissertation, U.C.L.A., 1966, pp. 3-4.

Nyberg, Verner R. "The Reliability of Essay Grading." Paper presented at the Sixth Annual Canadian Conference on Educational Research. Quebec City, 1968.

Nyberg, Verner R., and Nyberg, Adell M. Alberta Essay Scales: Models. Edmonton Alberta: Alberta Education, 1977.

Van Wagenen, M. J. Van Wagenen English Composition Scales. New York: World Book Company, 1923.

Willing, M. H. Willing Scale for Measuring Written Composition. Bloomington, Ill.: Public School Publishing, Co., 1922.

APPENDIX A

Rating Forms and Description of Variables

Form I

GRADE XII, 1964

EVALUATION OF LANGUAGE ESSAY FOR CORRECTNESS IN WRITTEN LANGUAGE

SCORE SHEET "A" (Deductions)

DEDUCT marks according to the score card														Enter total of deductions here
Spelling	3	3	3	3	3	3	3	3	3	3	3	3	3	
Punctuation	2	2	2	2	2	2	2	2	2	2	2	2	2	
Word Usage		2	2	2	2	2	2	2	2	2	2	2	2	
Grammar			3	3	3	3	3	3	3	3	3	3	3	
Sentence Errors	3	3	3	3	3	3	3	3	3	3	3	3	3	
Form							2	2	2	2	2	2	2	
													TOTAL DEDUCTIONS — (Maximum 125)	

Form II

GRADE XII -- 1964

INFORMATION RE SCORE SHEET A

For checking correctness in Written Language

DEDUCT marks.

1. SPELLING (including the apostrophe) _____

2. PUNCTUATION _____
   (a) Capitals.
   (b) Sentence endings.
   (c) Comma:
      (i) To separate items in addresses and dates.
      (ii) To separate words, phrases, or clauses in a series.
      (iii) To set off non-restrictive modifiers.
      (iv) To set off a phrase or clause at the beginning of a sentence.
      (v) To set off direct address and appositives.
      (iv) To set off words introducing a direct quotation.
   (d) Semicolon:
      (i) To separate principal clauses when the conjunction is omitted.
      (ii) Before a conjunction when principal clauses are broken by punctuation.
   (e) Quotation marks, single and double, for direct narration and quotations.
   (f) Colon, before a long quotation or a long series.

3. WORD USAGE _____
   (a) Impropriety—improper use of words.
   (b) Repetition.
   (c) Slang.
   (d) Redundancy.

4. GRAMMAR _____
   (a) Use of adjective for adverb.
   (b) Wrong use of prepositions and conjunctions, e.g. like and as.
   (c) Lack of agreement of subject and predicate.
   (d) Confusing past tense and perfect participle.
   (e) Lack of agreement of pronoun with antecedent.
   (f) Case errors.
   (g) Double negative.

5. SENTENCE ERRORS _____
   (a) Run-on sentences; excessive use of and and so.
   (b) Incomplete sentences.
   (c) Misplaced modifiers.
   (d) Indefinite reference of pronoun, participle, or gerund.
   (e) Tense sequence.
   (f) Error in use of parallel structure.
   (g) Changing the point of view, e.g., from present to past, active to passive.

6. FORM _____
   (a) Spacing on the page, and title.
   (b) General neatness.
   (c) Indentation of paragraphs, to indicate different speakers as well as divisions of subject matter.

Form III

GRADE XII, 1964.

Score Sheet B (Additions)—Evaluation of Language on the Essay

	Maximum	Student's Score
1. BASIC MARK	125	125
2. MATERIAL — 25 marks		
Significance	10	
Relevance	10	
Originality	5	
3. ORGANIZATION — 30 marks		
Plan	5	
Relationship of plan to essay	5	
Introduction	5	
Order:	5	
Logical sequence of developmental paragraphs Effective transition General coherence		
Emphasis	5	
Conclusion	5	
4. DICTION — 20 marks		
Exactness and vividness of nouns, verbs, adjectives, adverbs	10	
Figures of speech, comparisons, illustrations	5	
Breadth of vocabulary	5	
5. SENTENCE ORGANIZATION — 35 marks		
Variety:		
Structure	15	
Beginnings	5	
Economy	15	
6. TOTAL IMPRESSION — 15 marks	15	
Total		
Penalty for deficiency in length		
Net score for Score Sheet B		
Deduct total from Score Sheet A		
FINAL SCORE		

Form iv

GRADE XII, 1964

Notes for Sub-Examiners on the use of Section B

MATERIAL

Significance

Content should be worth writing. This may involve being convincing, i.e making thoughts sound genuine and situations appear real. Even good nonsense, if sustained, can be significant.

Relevance

All material included should be relevant to the topic.

One or two relevant points may be developed at length, or several developed in a more limited fashion.

Originality

Material should obviously have been drawn from personal experience.

ORGANIZATION

Plan

Any acceptable form.

Relationship of plan to essay

It should be evident that the plan has been followed in the writing of the essay.

Introduction

(a) arresting
(b) suggesting topic, clarifying where necessary, etc.
(c) setting mood
(d) raising a question (directly or indirectly)
(e) throwing interest ahead
(f) suggesting the pattern to follow (i.e the method of development)

Order

This implies
(a) evidence of a definite method of development
(b) use of appropriate transitional devices
(c) clarity of meaning in general (coherence)

Emphasis

(a) having a main point
(b) building to a climax (weightiest point last)
(c) highlighting major points and subordinating minor points

Conclusion

(a) memorable
(b) returning to introduction
(c) making title meaningful (making clear the connection between the title and the theme)
(d) providing original touch for emphatic close
(e) conveying a sense of finality (avoiding the mere "cut-off" ending)

DICTION

(a) particularly apt denotative and connotative words (specific nouns, lively verbs, vivid adjectives and adverbs)
(b) figures of speech: simile, metaphor, allegory, apostrophe, personification, hyperbole, metonymy, synecdoche, antithesis, irony

SENTENCE ORGANIZATION

Variety

Structure: Simple, complex, compound, compound-complex
Loose, periodic, balanced, parallel structure
Assertive, interrogative, exclamatory, imperative
Short, long

Beginning: Effective variety of beginnings.

Economy

Tightness and economy of structure through:
(a) use of verbals and verbal phrases (infinitive, gerund, participle)
(b) compound subjects and compound predicates
(c) balance and parallelism

Total Impression

Freshness of approach
Revelation of personality and expression of mood
Use of color, dialect
Characterization through dialogue
Humor or surprise twist (anti-climax, understatement, exaggeration, or overstatement, mock-seriousness, pretentious language for effect)
Effective use of devices such as examples, illustration, analogy, allusion, quotations, alliteration, assonance, rhythm
Effective touches of realism; sensory impressions aptly used